DOCUMENT RESUME

ED 330 719 TM 016 308

AUTHOR Bergstrom, Betty A.; Lunz, Mary E.
TITLE Confidence in Pass/Fail Decisions for Computer Adaptive and Paper and Pencil Examinations.
PUB DATE Apr 91
NOTE 23p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 3-7, 1991).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Adaptive Testing; Certification; Comparative Testing; *Computer Assisted Testing; Decision Making; Higher Education; *Item Response Theory; Licensing Examinations (Professions); Mastery Tests; *Medical Students; Medical Technologists; *Pass Fail Grading; Scoring; Test Length; Test Validity
IDENTIFIERS *Confidence Intervals (Statistics); *Paper and Pencil Tests; Rasch Model

ABSTRACT

The level of confidence in pass/fail decisions obtained with computer adaptive tests (CATs) was compared to decisions based on paper-and-pencil tests. Subjects included 645 medical technology students from 238 educational programs across the country. The tests used in this study constituted part of the subjects' review for the certification examination. Examinees took a variable length CAT and two fixed-length paper-and-pencil tests. The CAT, which was based on a pool of 726 Rasch calibrated items, was designed to determine whether an examinee's estimated ability level is above or below a pre-established pass/fail point. Short and long paper-and-pencil tests were used, containing 109 items and 189 items, respectively, selected from the CAT pool of 726 items. Only 600 examinees participated in the long paper-and-pencil test. The CAT was stopped when the examinee measure was 1.3 times the error of measure above or below the pass/fail point (one tailed 90% confidence interval) or when a maximum test length was reached. Results show that greater confidence in the accuracy of pass/fail decisions is obtained for more examinees when the CAT implements a 90% confidence stopping rule than with paper-and-pencil tests of comparable length. Using a confidence interval stopping rule with a CAT allowed each examinee to take a minimum number of items and insures that pass/fail decisions are made with a high level of confidence. Two tables and four figures are included. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED330719

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

BETTY BERGSTROM

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Confidence in Pass/Fail Decisions for Computer Adaptive and Paper and Pencil Examinations

Betty A. Bergstrom

Mary E. Lunz

American Society of Clinical Pathologists

BEST COPY AVAILABLE

Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL, April, 1991.

TM 016308

Confidence in Pass/Fail Decisions for

Computer Adaptive and Paper and Pencil Examinations

Abstract

The purpose of this study is to compare the level of confidence in pass/fail decisions obtained with computer adaptive tests and pencil and paper tests. Examinees took a variable length computer adaptive test and two fixed length pencil and paper tests. The computer adaptive test was stopped when the examinee measure was either 1.3 times the error of measure above or below the pass/fail point (one tailed 90% confidence interval) or when a maximum test length was reached. Results show that greater confidence in the accuracy of the pass/fail decisions is obtained for more examinees when the computer adaptive test implements a 90% confidence stopping rule than with paper and pencil tests of comparable test length.

Key words: computer adaptive test, confidence interval, stopping rule.

Confidence in Pass/Fail Decisions for

Computer Adaptive and Paper and Pencil Examinations

Licensure and Certifications examinations are often very long tests. Lengthy tests have been thought necessary in order to cover content areas and insure the reliability of pass/fail decisions. Certification and licensure examinations are high-stakes examinations where pass/fail decisions have great impact upon examinees. Confidence in the pass/fail decision is important because these examinations are designed to prevent incompetent individuals from practicing in the field.

Computer adaptive testing (CAT), the process of administering a test by computer so that only items of appropriate difficulty are administered to each examinee, has been shown to reduce test length without loss of precision (Weiss, 1983, 1985; Weiss and Kingsbury, 1984; McKinley and Reckase, 1980, 1984; Olsen, et. al., 1986). If it can be shown that pass/fail decisions made with CAT procedures are as reliable and valid as pass/fail decisions made with existing pencil and paper examinations, then computer adaptive testing offers advantages to both examinees and certification/licensure agencies with regard to shorter, more precise tests.

CAT Development

Concomitant developments in measurement theory and computer technology have made the use of computers for testing not only possible but practical and convenient (Reckase, 1989). To implement a CAT, Item Response Theory (IRT) is used to determine item calibrations and examinee ability measures (Lord and Novick, 1968; Wright and Stone, 1979). Each examinee takes an individualized test comprised of items chosen from a content valid item bank. After each

item is answered, the examinee ability measure is reestimated. If the examinee answers an item correctly, the estimation of his measure increases and the next item administered is more difficult; if the examinee answers an item incorrectly, the estimation of his ability is lowered and the next item administered is easier. This process utilizes an algorithm to estimate person abilities and to present the next best item given the test specifications for content coverage, test length and appropriate item difficulty. IRT based adaptive tests are greatly facilitated by a computer because of the computational requirements of the algorithm and because of the logistics of item and data management.

A basic consequence of Item Response Theory is the realization that items which are too easy or too difficult for an examinee contribute little information about that examinee's ability (Green et.al., 1984). But when items are targeted to the ability level of the examinee, the error of measure (SEM) is minimized so that test length can be reduced without loss of precision (Wright and Stone, 1979).

Stopping Rules

Computer adaptive tests can be either fixed or variable length depending upon the stopping rule implemented. If the CAT is terminated after a specified number of items, it is a fixed length test. Two slightly different stopping rules result in variable length tests. Examinees can be tested to a specified level of precision (SEM) or they can be tested to a specified level of confidence in a pass/fail decision. Regardless of the stopping rule implemented, the number of items required to pass or fail an examinee with the same degree of precision or confidence should be less for a CAT than the number of items required by a fixed length pencil and paper test.

SEM Stopping Rule When a specified standard error of measure (SEM) is used as the stopping rule, test length is variable, but the range of variability is limited because a given number of items will be required to decrease the SEM to the designated level. When the Rasch model is used, SEM can be calculated from the formula

$$SEM_n = [1/\sum_{i}^{L}(P_{ni}\ (1-P_{ni}))]^{1/2}$$

with P_{ni} being the probability of examinee$_n$ responding correctly to an item and $\sum_{i}^{L}$ being the summation of $P_{ni}\ (1-P_{ni})$ over the items encountered. Test lengths for specified levels of SEM can be approximated from the formula

$$Length = 1/(SEM^2[P(1-P)]).$$

by assuming equal probability (P) of correct response over all items for a given examinee (Wright and Stone, 1979). For example, if the probability of correct response ranges from .50 to .10 for different examinees, it would require approximately 64 to 177 items to reach a SEM of .25. When a SEM stopping rule is employed, all examinees are tested to the same level of precision regardless of the number of items required to reach the specified SEM.

Confidence Level Stopping Rule A one-tailed confidence level can be estimated by calculating the ratio of the distance of the examinee ability measure (B) from the pass/fail point (PF) divided by the standard

$$(B-PF)/SEM$$

error of measure (SEM) and comparing this ratio to the normal probability distribution table. When a specified level of confidence is used as the stopping rule, both the distance from the pass/fail point and the SEM influence test length. Using a confidence level stopping rule means that a

more precise estimate of ability is obtained for examinees whose estimated ability is near the pass/fail point, because they take longer tests. It also means that while abilities of very high ability and very low ability examinees are estimated less precisely, a high level of confidence in their pass/fail decision is obtained and they take shorter tests.

Use of a confidence level stopping rule requires that a maximum test length be set because the required number of items to reach the specified level of confidence increases greatly as the examinee ability measure gets very near the pass/fail point. If, for example, an examinee ability measure is only .10 logits from the pass/fail point, that examinee would have to take over 600 perfectly targeted items to reach a confidence level of 90%, assuming his ability measure remained stable.

Purpose

The purpose of this study is to compare the level of confidence in pass/fail decisions obtained with computer adaptive tests and pencil and paper tests. This assumes the tests are equated, the same pass/fail point is set for all tests, and the CAT item selection algorithm fulfills test specifications.

Method

Subjects

Subjects were 645 medical technology students, from 238 educational programs across the country. To participate, students had to be eligible to take the summer administration of the related certification examination. Students who participated were serious about the experience since it was part of their review for the certification examination.

Procedures

Examinees took a variable length computer adaptive test, a short paper and pencil test (109 items) and a long paper and pencil test (189 items).

Tests

Computer Adaptive. Each examinee took an individualized computer adaptive test selected from a pool of 726 Rasch calibrated items. The test was designed as an adaptive mastery test (Weiss and Kingsbury, 1984) to determine whether an examinees estimated ability level is above or below a pre-established pass/fail point.

The test was administered by the CAT Administrator (Gershon, 1989) which implemented the PROX version of the maximum likelihood method of item selection (Wright and Stone, 1979) in the adaptive algorithm. The calibrated item pool contained items from six subsets. Table 1 provides a summary of the item pool distribution. Content coverage was designed to be comparable to the test specification for the paper and pencil examinations so a content balancing algorithm of the type described by Kingsbury and Zara (1989) was included in the item selection algorithm. Subset order was selected randomly by the computer algorithm. Maurelli and Weiss (1981) found subtest order to have no effect on the psychometric properties of an existent achievement test battery.

Items were targeted to the current ability measure of the examinee and were chosen at random from unused items within .10 logits of the targeted item difficulty within the selected subset. Kingsbury and Houser (1990) have shown that an adaptive testing procedure which provides maximum information about the examinee's ability will provide a more clear indication that the examinee is above or below the pass/fail point than a test which peaks

information at the pass/fail point. While the examinee considered the item presented, the computer selected two items within the specified content area, one which would yield maximum information should the current item be answered correctly and another which would yield maximum information should the current item be answered incorrectly. This procedure insured that there was no lag time before the next item was presented.

The stopping rule employed was based on confidence in the pass/fail decision. The test was stopped when the examinee measure was either 1.3 times the error of measure above the pass/fail point (a clear pass--one tailed 90% confidence level) or 1.3 times the error of measure below the pass/fail point (a clear fail) or when a maximum test length was reached. Minimum test length was 50 items; maximum test length was 240 items. Confidence levels were also analyzed using 100 items as the maximum test length.

Short Paper and Pencil. The short paper and pencil (PAP) test was given during the same week as the computer adaptive test. It contained 109 items selected from the computer adaptive test pool of 726 items. It was analyzed with BIGSCALE (Wright, Linacre and Schultz, 1990), a Rasch program for item calibration and examinee ability estimation, and equated to the computer adaptive test with common item equating (Wright and Stone, 1979).

Long Paper and Pencil. The long paper and pencil (PAP) test was administered approximately two months after the CAT. Six hundred (600) of the 645 examinees participated. Test length was 189 items and there were no items common to either the CAT or the short paper and pencil test. The long pencil and paper test was analyzed with BIGSCALE and equated to the other two tests with common person equating (Wright and Stone, 1979).

Results

Summary statistics for the four tests are presented in Table 2. There are no significant differences in mean ability measures (α=.05) across tests. All examinees took 109 items on the short PAP and 189 items on the long PAP. Mean test length for the CAT (MAX 240) was 85 items and mean test length for the CAT (MAX 100) was 66 items.

Figure 1 shows the confidence levels for decisions made with the computer adaptive tests, (stopping the test at a maximum of 240 items and stopping the test at a maximum of 100 items) compared to the confidence levels for decisions made with the paper and pencil tests (long and short). Ninety percent of examinees passed or failed the long CAT (max test length 240 items, mean test length 86 items) with 90% confidence in the accuracy of the decision while only 72% of examinees passed or failed the long paper and pencil test (test length 189 items) with 90% confidence in the accuracy of the decisions. When the CAT results were calculated using 100 items as the maximum test length (mean 67 items) 90% confidence in the accuracy of the pass/fail decisions was obtained for 73% of examinees compared to 53% of the examinees on the short paper and pencil test (test length 109 items).

The Appendix contains three examples of examinee records from the computer adaptive test. Examinee A clearly failed the CAT (> 99% confidence in the pass/fail decision). Examinee B clearly passed the CAT (>99% confidence in the pass/fail decision). Testing stopped at the minimum number of items (50) for both of these examinees. Examinee C is an example of an examinee whose ability measure is very close to the pass/fail point of .15 logits. Test length continued to the maximum test length of 240 items.

Discussion

Stopping Rules

Choosing a stopping rule depends upon the purpose of the test. A confidence level stopping rule is especially useful for certification tests where the main concern is the integrity of the pass/fail decision. Examinees whose estimated ability measure is near the pass/fail point take longer tests resulting in a more precise estimate of their ability than examinees whose ability measures are further from the pass/fail point. Examinees who clearly pass or fail the examination (90% or more confidence in the accuracy of the decision) require shorter, less precise tests because the confidence in the decision is very high.

Certification examinations require that a decision be made about an examinee relative to their ability to practice in a certain field as determined by a criterion referenced test. Definitive decisions are highly desirable even though it is common for 30% of the population to have ability measures within the error of measurement that surrounds the pass/fail point. The results of this study show that CAT pass/fail decisions are made with comparable or greater confidence in their accuracy than pass/fail decisions made with paper and pencil tests. This is possible because the CAT is targeted to the ability level of the examinee.

Maximum Test Length

Maximum test length can be set to insure that confidence levels in pass/fail decisions are equivalent to, or exceed, confidence levels obtained in paper and pencil certification tests. However, the map for Examinee C (Appendix) illustrates the need for maximum test length to be set at a reasonable limit. Definitive pass/fail decisions cannot be made for examinees

whose ability measures are very close to the pass/fail point with paper and pencil or computer adaptive tests. Giving an excessive number of items on the computer adaptive test contributes little additional information about the examinee. The map of Examinee C also shows that as the number of items increases, the items may become less well targeted because the most appropriate items have already been presented.

Conclusion

In order for computer adaptive testing to be implemented by certification agencies, benefits need to be confirmed by basic research studies. This study contributes to knowledge about the implications of using a computer adaptive test with a confidence level stopping rule for certification. Implementation of computer adaptive testing for certification can reduce test length and improve confidence in the accuracy of the pass/fail decisions. Using a confidence interval stopping rule with a computer adaptive test allows each examinee to take a minimum number of items and insures that pass/fail decisions are made with a high level of confidence. Thus, computer adaptive testing appears to be a potentially useful method of administration for certification examinations.

Reference

Green, B.F., Bock, R.D., Humphreys, L.G., Linn, R.L. and Reckase, M.D. (1984). Technical guidelines for assessing computerized adaptive tests. Journal of Educational Measurement, 21, 4, 347-360.

Kingsbury, G.G. and Houser, R.L. (1990). Assessing the utility of item response models: Computerized adaptive testing. Paper presented to the annual meeting of the National Council on Measurement in Education, Boston, MA.

Kingsbury, G.G. and Zara, A.R. (1989). Procedures for selecting items for computerized adaptive tests. Applied Measurement in Education, 2, 4, 359-375.

Lord, F.M. and Novick, M.R. (1968). Statistical Theories of Mental Test Score. Reading, Mass: Addison-Wesley.

McKinley, R.L. and Reckase, M.D. (1980). Computer applications to ability testing. Association for Educational Data Systems Journal, 13, 193-203.

McKinley, R.L. and Reckase, M.D. (1984). Impementing an adaptive testing progr[illegible] in an instructional program environment. Paper presented at the meeting of the American Educational Research Association, New Orleans.

Maurelli, V.A. and Weiss, D.J. (1983). Factors influencing the psychometric characteristics of an adaptive testing strategy for test batteries (Research Report 81-4). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, Computerized Adaptive Testing laboratory.

Olsen, J.B., Maynes, D.D., Slawson, D. and Ho, K. (1986). Comparison and equating of paper-administered, computer-administered and computerized adaptive tests of achievement. Paper presented at the American Educational Research Association Meeting, San Francisco, CA.

Reckase, M.D. (1989). Adaptive testing: The evolution of a good idea. Educational Measurement: Issues and Practice, 8, 3.

Weiss, D.J. (1983). New horizons in testing: Latent trait test theory and computerized adaptive testing. Academic Press Inc., New York.

Weiss, D.J. (1985). Final Report: Computerized Adaptive Measurement of Achievement and Ability (Project NR150-433, N00014-79-C0172 Minneapolis, MN: University of Minnesota.

Weiss, D.J. and Kingsbury, G.G. (1984). Application of computerized adaptive testing to educational problems. Journal of Educational Measurement, 21, 4, 361-375.

Wright, B.D., Linacre, J.M. and Schultz, M. (1989). BIGSCALE (Computer Program). Chicago: University of Chicago.

Wright, B.D. and Stone, M.H. (1979). Best Test Design. Chicago: MESA Press.

Appendix

Map Key

A. Summary Statistics

1. Total Test/Subtest
   - T - total
   - M - microbiology
   - BB - blood banking
   - C - chemistry
   - H - hematology
   - BF - body fluids
   - I - immunology

B. Number of items

C. Number of items correct

D. Number of items incorrect

E. Ability measure

F. Error of measure

G. Average item difficulty

H. Sum of the squares (item difficulty)

I. Item number

J. Subtest identifier

K. Item difficulty

L. Response selected

M. Right/wrong 0-incorrect 1-correct

N. Current estimated ability measure

O. Current estimated error

P. Time (in seconds)

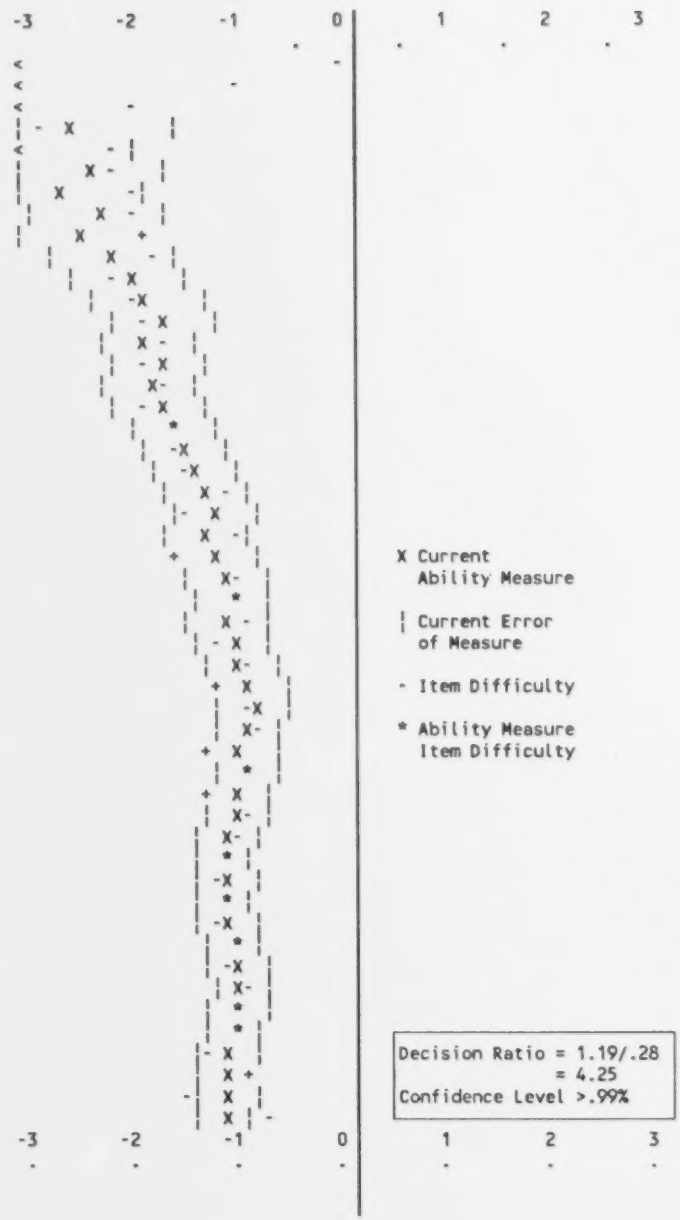
Pass/fail point (.15 logits)

Map for Examinee A

Clear Fail

A	B	C	D	E	F	G	H
T	50	28	22	-1.04	0.28	-1.28	94.7
M	10	5	5	-0.96	0.63	-0.96	9.4
BB	10	6	4	-0.54	0.65	-0.95	9.3
C	10	8	2	-0.29	0.79	-1.68	28.5
H	10	4	6	-2.13	0.65	-1.72	35.1
BF	5	4	1	0.23	1.12	-1.16	7.0
I	5	1	4	-2.36	1.12	-0.98	5.3

	I	J	K	L	M	N	O	P
. 1	520	H	0.07	4	0	-9.99	-9.99	63
2	481	H	-0.94	4	0	-9.99	-9.99	10
3	458	H	-1.95	4	0	-9.99	-9.99	17
4	455	H	-2.80	1	1	-2.50	1.15	16
5	456	H	-2.15	1	0	-2.94	1.12	17
6	457	H	-2.13	1	1	-2.34	0.87	6
7	459	H	-1.91	4	0	-2.60	0.84	32
8	460	H	-1.87	4	1	-2.22	0.73	6
9	461	H	-1.83	2	0	-2.42	0.71	14
10	462	H	-1.72	4	1	-2.13	0.65	20
11	292	C	-2.10	1	1	-1.94	0.61	9
12	293	C	-1.87	1	1	-1.77	0.58	26
13	296	C	-1.76	1	1	-1.61	0.56	20
14	300	C	-1.57	1	0	-1.75	0.53	17
15	298	C	-1.76	3	1	-1.62	0.52	36
16	301	C	-1.57	1	0	-1.74	0.50	20
17	297	C	-1.76	1	1	-1.62	0.49	5
18	302	C	-1.55	4	1	-1.51	0.47	8
19	306	C	-1.46	4	1	-1.40	0.46	11
20	309	C	-1.37	4	1	-1.29	0.46	54
21	595	BF	-1.01	1	1	-1.18	0.45	15
22	594	BF	-1.37	2	1	-1.09	0.44	6
23	596	BF	-0.95	2	0	-1.18	0.43	7
24	593	BF	-1.55	4	1	-1.11	0.42	2
25	597	BF	-0.91	3	1	-1.02	0.42	5
26	170	BB	-0.94	2	1	-0.93	0.41	98
27	171	BB	-0.84	3	0	-1.01	0.40	17
28	169	BB	-1.08	3	1	-0.94	0.39	12
29	172	BB	-0.84	2	1	-0.86	0.39	12
30	168	BB	-1.08	1	1	-0.79	0.39	24
31	173	BB	-0.79	4	1	-0.72	0.38	12
32	179	BB	-0.71	2	0	-0.79	0.37	21
33	167	BB	-1.20	4	0	-0.87	0.36	12
34	174	BB	-0.79	4	1	-0.81	0.36	11
35	166	BB	-1.23	1	0	-0.88	0.35	8
36	34	M	-0.82	1	0	-0.94	0.34	20
37	32	M	-0.94	4	0	-1.00	0.33	32
38	27	M	-1.01	2	0	-1.05	0.33	14
39	20	M	-1.08	1	1	-1.00	0.33	20
40	28	M	-0.99	4	0	-1.05	0.32	13
41	19	M	-1.13	1	1	-1.00	0.32	28
42	29	M	-0.94	3	1	-0.95	0.31	9
43	26	M	-1.01	3	1	-0.91	0.31	14
44	35	M	-0.79	2	1	-0.86	0.31	11
45	33	M	-0.93	2	0	-0.90	0.30	13
46	668	I	-0.88	2	0	-0.95	0.30	32
47	667	I	-1.24	3	0	-1.00	0.30	13
48	669	I	-0.76	2	0	-1.04	0.29	15
49	666	I	-1.42	2	1	-1.00	0.29	16
50	670	I	-0.59	3	0	-1.04	0.28	13

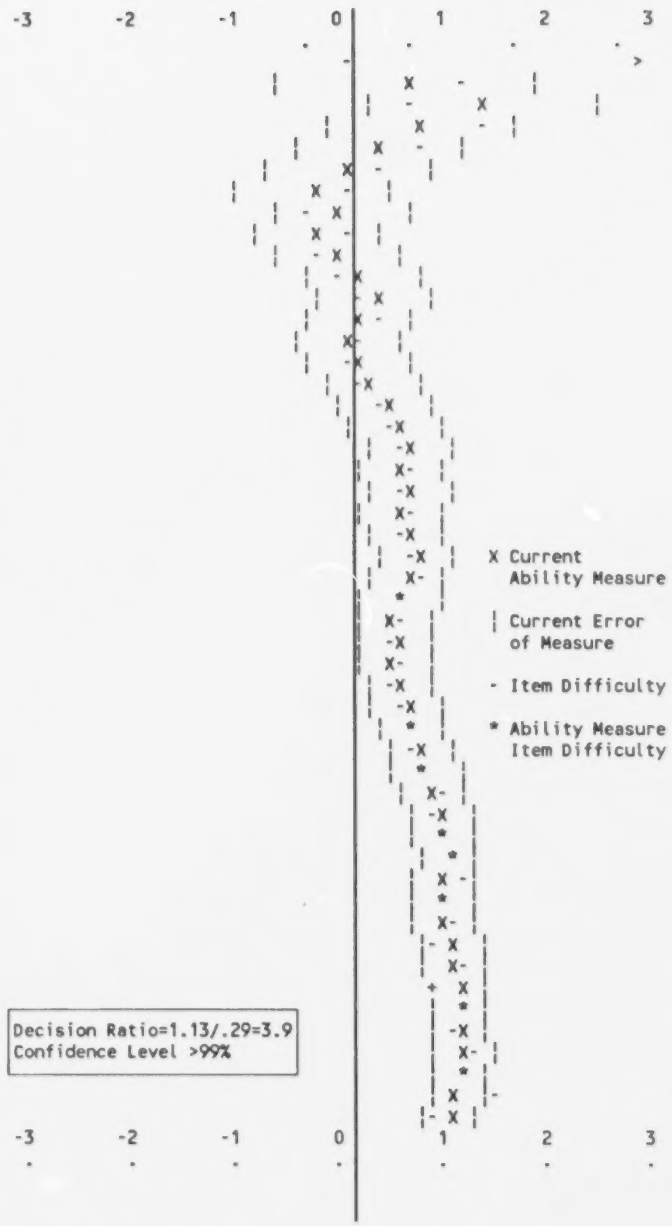


Map for Examinee B

Clear Pass

A	B	C	D	E	F	G	H
T	50	31	19	1.28	0.29	0.79	39.9
M	10	5	5	0.71	0.63	0.71	5.1
BB	10	7	3	1.28	0.69	0.43	2.4
C	10	8	2	2.52	0.79	1.13	13.0
H	10	4	6	0.12	0.65	0.53	5.8
BF	5	2	3	0.98	0.91	1.39	9.8
I	5	5	0	9.99	9.99	0.86	3.8

I	J	K	L	M	N	O	P	
.	1	532	H	0.22	1	1	9.99	9.99	81
	2	582	H	1.31	3	0	0.76	1.41	5
	3	566	H	0.76	4	1	1.46	1.22	10
	4	587	H	1.49	4	0	0.94	1.00	54
	5	573	H	0.92	2	0	0.53	0.91	15
	6	548	H	0.51	3	0	0.18	0.87	26
	7	527	H	0.17	2	0	-0.15	0.84	38
	8	510	H	-0.16	4	1	0.14	0.73	7
	9	526	H	0.17	2	0	-0.09	0.71	22
	10	515	H	-0.10	2	1	0.12	0.65	29
	11	230	BB	0.10	3	1	0.31	0.61	7
	12	239	BB	0.31	3	1	0.47	0.58	16
	13	251	BB	0.51	1	0	0.32	0.56	28
	14	242	BB	0.31	4	0	0.18	0.54	81
	15	234	BB	0.17	3	1	0.31	0.52	31
	16	243	BB	0.35	1	1	0.44	0.50	37
	17	250	BB	0.46	2	1	0.56	0.49	20
	18	253	BB	0.60	2	1	0.67	0.47	8
	19	259	BB	0.74	1	1	0.78	0.46	9
	20	262	BB	0.79	3	0	0.68	0.45	27
	21	115	M	0.68	4	1	0.78	0.44	20
	22	119	M	0.78	4	0	0.69	0.43	34
	23	114	M	0.68	2	1	0.77	0.42	13
	24	120	M	0.79	1	1	0.86	0.41	12
	25	124	M	0.86	3	0	0.78	0.40	16
	26	118	M	0.74	2	0	0.70	0.39	11
	27	113	M	0.67	4	0	0.62	0.39	17
	28	109	M	0.61	3	1	0.69	0.38	19
	29	116	M	0.72	1	0	0.63	0.37	12
	30	108	M	0.60	4	1	0.69	0.37	36
	31	703	I	0.66	1	1	0.76	0.36	20
	32	705	I	0.78	3	1	0.82	0.36	29
	33	707	I	0.83	2	1	0.88	0.35	29
	34	710	I	0.94	1	1	0.94	0.35	38
	35	711	I	1.09	3	1	1.01	0.35	24
	36	431	C	1.00	1	1	1.06	0.34	5
	37	434	C	1.09	3	1	1.12	0.34	8
	38	435	C	1.17	4	1	1.18	0.34	45
	39	437	C	1.25	1	0	1.13	0.33	14
	40	433	C	1.06	1	0	1.07	0.32	35
	41	436	C	1.17	2	1	1.12	0.32	9
	42	432	C	1.03	1	1	1.17	0.32	18
	43	438	C	1.28	1	1	1.22	0.32	10
	44	430	C	0.98	3	1	1.27	0.31	12
	45	439	C	1.31	1	1	1.31	0.31	26
	46	655	BF	1.25	2	1	1.36	0.31	37
	47	657	BF	1.50	3	1	1.41	0.31	28
	48	656	BF	1.36	1	0	1.36	0.30	61
	49	658	BF	1.70	2	0	1.32	0.30	17
	50	654	BF	1.13	4	0	1.28	0.29	11

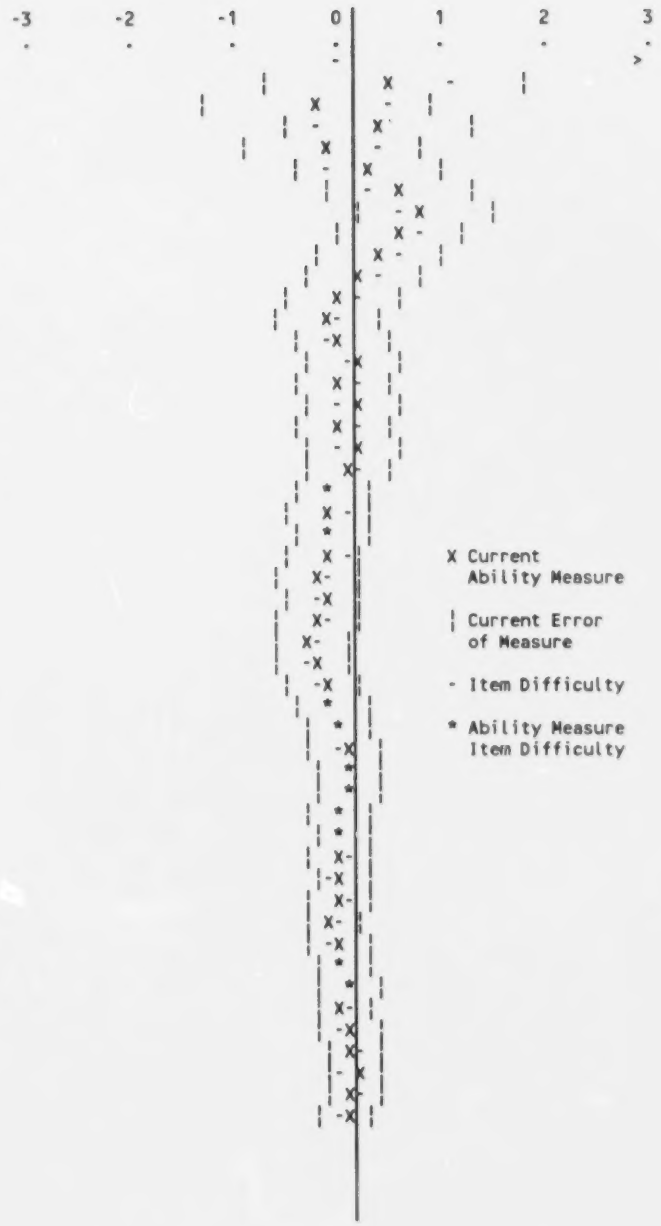


Map for Examinee C

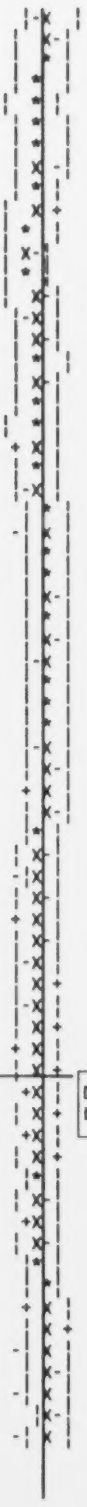
Undecided

A	B	C	D	E	F	G	H
T	240	123	117	0.19	0.13	0.14	20.7
M	50	28	22	0.41	0.28	0.17	5.8
BB	50	25	25	0.20	0.28	0.20	4.7
C	46	24	22	0.18	0.30	0.10	2.5
H	46	23	23	0.09	0.29	0.09	2.9
BF	23	10	13	-0.13	0.42	0.13	1.4
I	25	13	12	0.23	0.40	0.15	3.4

	I	J	K	L	M	N	O	P
1	691	I	0.13	1	1	9.99	9.99	7
2	713	I	1.17	4	0	0.65	1.41	47
3	702	I	0.60	3	0	-0.06	1.22	24
4	686	I	-0.06	4	1	0.46	1.00	26
5	699	I	0.46	1	0	0.05	0.91	21
6	78	M	0.03	4	1	0.39	0.82	19
7	99	M	0.40	1	1	0.68	0.76	19
8	114	M	0.68	2	1	0.94	0.73	22
9	126	M	0.95	2	0	0.71	0.67	39
10	115	M	0.68	3	0	0.50	0.63	17
11	104	M	0.49	1	0	0.32	0.61	37
12	94	M	0.31	1	0	0.15	0.59	25
13	83	M	0.15	1	0	-0.01	0.57	17
14	76	M	-0.05	3	1	0.14	0.54	34
15	84	M	0.17	4	1	0.27	0.52	13
16	392	C	0.25	4	0	0.15	0.50	21
17	380	C	0.13	2	1	0.26	0.49	9
18	393	C	0.30	3	0	0.15	0.47	25
19	379	C	0.07	1	1	0.26	0.46	11
20	394	C	0.30	4	0	0.16	0.45	19
21	378	C	0.01	3	0	0.05	0.44	11
22	381	C	0.17	2	0	-0.03	0.43	7
23	375	C	-0.05	4	1	0.05	0.42	17
24	382	C	0.19	4	0	-0.02	0.41	21
25	374	C	-0.05	1	0	-0.11	0.41	13
26	514	H	-0.10	1	1	-0.03	0.40	31
27	518	H	0.01	3	0	-0.10	0.39	34
28	513	H	0.11	1	0	-0.18	0.39	26
29	507	H	-0.20	4	1	-0.11	0.38	18
30	515	H	-0.10	2	1	-0.04	0.37	36
31	519	H	0.01	2	1	0.03	0.36	50
32	520	H	0.07	2	1	0.09	0.35	18
33	523	H	0.13	4	1	0.16	0.35	12
34	526	H	0.17	3	1	0.21	0.34	8
35	532	H	0.22	2	0	0.16	0.34	16
36	627	BF	0.13	3	0	0.10	0.33	18
37	624	BF	0.07	3	1	0.15	0.33	9
38	628	BF	0.19	3	0	0.10	0.32	49
39	623	BF	0.01	1	1	0.15	0.32	6
40	629	BF	0.20	1	0	0.10	0.32	27
41	230	BB	0.10	1	0	0.05	0.31	31
42	225	BB	0.03	1	1	0.10	0.31	6
43	231	BB	0.13	4	1	0.15	0.31	40
44	234	BB	0.17	3	1	0.19	0.30	20
45	235	BB	0.24	1	0	0.15	0.30	31
46	233	BB	0.13	4	1	0.19	0.29	22
47	236	BB	0.25	3	1	0.24	0.29	17
48	232	BB	0.13	4	1	0.28	0.29	21
49	238	BB	0.30	2	0	0.24	0.29	7
50	229	BB	0.10	1	0	0.19	0.28	14



51	626	BF	0.13	2	1	0.23	0.28	9	F
52	631	BF	0.25	3	0	0.19	0.28	10	F
53	531	H	0.19	2	0	0.16	0.27	28	F
54	525	H	0.13	3	0	0.12	0.27	96	F
55	522	H	0.08	2	0	0.08	0.27	16	F
56	524	H	0.13	2	1	0.12	0.27	12	F
57	228	BB	0.08	2	1	0.15	0.26	44	F
58	237	BB	0.25	2	0	0.12	0.26	18	F
59	227	BB	0.08	1	0	0.09	0.26	12	F
60	239	BB	0.31	2	0	0.06	0.26	32	F
61	687	I	0.00	1	0	0.02	0.26	20	F
62	688	I	0.07	4	0	-0.01	0.26	49	F
63	376	C	0.00	2	1	0.02	0.25	18	F
64	383	C	0.19	2	1	0.06	0.25	55	F
65	377	C	0.00	1	1	0.09	0.25	14	F
66	384	C	0.19	2	1	0.12	0.25	19	F
67	80	M	0.07	4	1	0.15	0.24	27	F
68	85	M	0.17	3	0	0.12	0.24	28	F
69	79	M	0.07	2	0	0.09	0.24	13	F
70	81	M	0.13	2	0	0.06	0.24	42	F
71	685	I	-0.10	1	1	0.09	0.24	18	F
72	690	I	0.10	1	1	0.11	0.24	25	F
73	373	C	-0.05	1	1	0.14	0.23	14	F
74	385	C	0.22	2	1	0.17	0.23	15	F
75	372	C	-0.06	4	1	0.19	0.23	9	F
76	386	C	0.22	2	1	0.22	0.23	23	F
77	530	H	0.19	4	1	0.24	0.23	12	F
78	533	H	0.25	2	0	0.22	0.23	74	F
79	529	H	0.18	1	0	0.19	0.23	24	F
80	534	H	0.30	1	0	0.17	0.22	32	F
81	82	M	0.13	1	1	0.19	0.22	45	F
82	87	M	0.22	4	1	0.22	0.22	13	F
83	86	M	0.19	4	0	0.19	0.22	77	F
84	88	M	0.24	3	1	0.22	0.22	35	F
85	226	BB	0.08	2	0	0.19	0.22	62	F
86	240	BB	0.31	4	1	0.22	0.22	16	F
87	224	BB	0.00	2	0	0.19	0.21	30	F
88	241	BB	0.31	1	0	0.17	0.21	45	F
89	625	BF	0.13	4	0	0.15	0.21	38	F
90	630	BF	0.22	4	0	0.13	0.21	29	F
91	371	C	-0.06	2	1	0.15	0.21	14	F
92	387	C	0.22	4	0	0.12	0.21	43	F
93	370	C	-0.11	1	0	0.10	0.21	7	F
94	388	C	0.24	1	0	0.08	0.21	19	F
95	622	BF	-0.05	4	1	0.10	0.21	22	F
96	632	BF	0.25	3	0	0.08	0.20	18	F
97	77	M	0.01	3	1	0.10	0.20	8	F
98	89	M	0.25	4	1	0.12	0.20	58	F
99	75	M	-0.06	1	0	0.10	0.20	22	F
100	90	M	0.25	4	1	0.12	0.20	50	F
101	223	BB	-0.01	2	1	0.14	0.20	22	F
102	242	BB	0.31	4	0	0.12	0.20	95	F
103	222	BB	-0.01	3	1	0.14	0.20	15	F
104	243	BB	0.35	3	0	0.12	0.20	30	F
105	521	H	0.08	2	1	0.14	0.20	11	F
106	527	H	0.17	1	0	0.12	0.19	19	F
107	517	H	-0.05	2	1	0.14	0.19	45	F
108	528	H	0.18	3	0	0.12	0.19	58	F
109	689	I	0.08	4	1	0.14	0.19	15	F
110	692	I	0.22	1	1	0.16	0.19	28	F
111	621	BF	-0.05	4	1	0.18	0.19	11	F
112	633	BF	0.36	2	1	0.20	0.19	24	F
113	684	I	-0.11	3	1	0.21	0.19	20	F
114	693	I	0.31	3	0	0.19	0.19	25	F
115	74	M	-0.06	2	1	0.21	0.19	33	F
116	91	M	0.27	4	1	0.23	0.19	23	F
117	73	M	-0.11	4	0	0.21	0.18	9	F



Decision Ration=.03/.20=.15
Confidence Level ~ 56%

118	92	M	0.30	3	0	0.19	0.18	47	F
119	369	C	-0.11	1	1	0.20	0.18	31	F
120	389	C	0.24	3	1	0.22	0.18	14	F
121	368	C	-0.11	4	0	0.20	0.18	14	F
122	390	C	0.25	2	1	0.22	0.18	21	F
123	221	BB	-0.01	4	0	0.20	0.18	8	F
124	244	BB	0.36	1	1	0.22	0.18	19	F
125	220	BB	-0.01	1	1	0.23	0.18	20	F
126	245	BB	0.41	1	0	0.22	0.18	24	F
127	516	H	-0.05	2	1	0.23	0.18	16	F
128	535	H	0.30	3	1	0.25	0.18	18	F
129	512	H	-0.11	3	0	0.23	0.18	13	F
130	536	H	0.30	2	0	0.21	0.18	53	F
131	683	I	-0.11	4	0	0.20	0.17	6	F
132	694	I	0.31	4	0	0.18	0.17	26	F
133	620	BF	-0.06	3	0	0.17	0.17	28	F
134	634	BF	0.36	4	0	0.15	0.17	44	F
135	72	M	-0.15	4	1	0.17	0.17	14	F
136	93	M	0.31	2	0	0.15	0.17	35	F
137	71	M	-0.15	4	1	0.16	0.17	16	F
138	95	M	0.35	1	0	0.15	0.17	16	F
139	219	BB	-0.05	1	1	0.16	0.17	47	F
140	246	BB	0.42	3	1	0.18	0.17	20	F
141	218	BB	-0.05	2	1	0.19	0.17	33	F
142	247	BB	0.42	1	0	0.18	0.17	23	F
143	511	H	-0.11	4	1	0.19	0.17	26	F
144	537	H	0.31	2	1	0.21	0.17	39	F
145	510	H	-0.16	4	1	0.22	0.17	11	F
146	538	H	0.35	4	1	0.23	0.17	26	F
147	367	C	-0.11	2	1	0.24	0.17	24	F
148	391	C	0.25	4	0	0.23	0.16	16	F
149	366	C	-0.15	1	1	0.24	0.16	19	F
150	395	C	0.30	4	0	0.23	0.16	10	F
151	217	BB	-0.05	2	1	0.24	0.16	13	F
152	248	BB	0.43	3	0	0.23	0.16	33	F
153	216	BB	-0.06	1	0	0.21	0.16	52	F
154	249	BB	0.43	4	1	0.23	0.16	26	F
155	619	BF	-0.06	2	1	0.24	0.16	19	F
156	635	BF	0.37	4	1	0.25	0.16	36	F
157	365	C	-0.15	1	1	0.26	0.16	36	F
158	396	C	0.30	2	0	0.25	0.16	39	F
159	364	C	-0.16	3	1	0.26	0.16	10	F
160	397	C	0.31	3	0	0.25	0.16	26	F
161	682	I	-0.16	1	1	0.26	0.16	24	F
162	695	I	0.36	1	0	0.25	0.16	6	F
163	70	M	-0.15	2	0	0.23	0.16	47	F
164	96	M	0.36	1	0	0.22	0.16	16	F
165	69	M	-0.16	4	1	0.23	0.16	21	F
166	97	M	0.36	1	1	0.24	0.16	22	F
167	509	H	-0.16	2	0	0.23	0.15	81	F
168	539	H	0.36	4	0	0.22	0.15	20	F
169	508	H	-0.16	1	0	0.20	0.15	8	F
170	540	H	0.40	4	1	0.22	0.15	43	F
171	618	BF	-0.20	1	0	0.20	0.15	18	F
172	636	BF	0.41	3	1	0.21	0.15	25	F
173	68	M	-0.16	3	1	0.22	0.15	11	F
174	98	M	0.37	3	1	0.24	0.15	20	F
175	67	M	-0.16	3	0	0.22	0.15	12	F
176	100	M	0.43	2	0	0.21	0.15	17	F
177	506	H	-0.22	4	1	0.22	0.15	8	F
178	541	H	0.40	4	0	0.21	0.15	12	F
179	505	H	-0.24	2	1	0.22	0.15	38	F
180	542	H	0.42	4	1	0.23	0.15	19	F
181	363	C	-0.16	2	1	0.24	0.15	31	F
182	398	C	0.36	3	0	0.23	0.15	27	F
183	362	C	-0.20	2	1	0.24	0.15	28	F
184	399	C	0.40	3	0	0.23	0.15	29	F

185	681	I	-0.16	2	0	0.22	0.15	30	F
186	696	I	0.37	4	0	0.21	0.15	16	F
187	215	BB	-0.06	4	0	0.19	0.15	32	F
188	250	BB	0.46	2	1	0.21	0.15	13	F
189	214	BB	-0.06	3	1	0.22	0.15	31	F
190	251	BB	0.51	3	0	0.21	0.15	32	F
191	680	I	-0.22	1	1	0.21	0.14	35	F
192	697	I	0.40	3	1	0.23	0.14	10	F
193	617	BF	-0.28	4	0	0.21	0.14	14	F
194	637	BF	0.41	3	0	0.20	0.14	50	F
195	213	BB	-0.06	3	0	0.19	0.14	8	F
196	252	BB	0.54	1	1	0.20	0.14	37	F
197	212	BB	-0.10	4	1	0.21	0.14	71	F
198	253	BB	0.60	1	0	0.20	0.14	19	F
199	504	H	-0.26	3	1	0.21	0.14	32	F
200	543	H	0.43	2	0	0.20	0.14	15	F
201	503	H	-0.28	3	0	0.19	0.14	32	F
202	544	H	0.46	3	0	0.18	0.14	21	F
203	66	M	-0.22	3	1	0.19	0.14	38	F
204	101	M	0.43	2	1	0.20	0.14	33	F
205	65	M	-0.22	2	1	0.21	0.14	22	F
206	102	M	0.48	3	0	0.20	0.14	15	F
207	361	C	-0.20	4	1	0.21	0.14	31	F
208	400	C	0.41	1	0	0.20	0.14	14	F
209	360	C	-0.21	2	1	0.21	0.14	7	F
210	401	C	0.42	2	0	0.20	0.14	20	F
211	64	M	-0.24	1	1	0.21	0.14	28	F
212	103	M	0.49	1	0	0.20	0.14	24	F
213	63	M	-0.26	3	1	0.21	0.14	22	F
214	105	M	0.54	1	0	0.20	0.14	15	F
215	616	BF	-0.28	2	1	0.20	0.14	16	F
216	638	BF	0.43	1	0	0.20	0.14	20	F
217	359	C	-0.22	4	0	0.18	0.14	20	F
218	402	C	0.43	3	0	0.18	0.14	37	F
219	358	C	-0.24	2	1	0.18	0.14	43	F
220	403	C	0.46	1	0	0.18	0.13	13	F
221	211	BB	-0.10	3	1	0.18	0.13	13	F
222	254	BB	0.60	2	0	0.18	0.13	33	F
223	210	BB	-0.10	4	1	0.18	0.13	13	F
224	255	BB	0.62	4	1	0.20	0.13	25	F
225	679	I	-0.34	4	1	0.20	0.13	8	F
226	698	I	0.40	4	1	0.21	0.13	33	F
227	502	H	-0.28	2	1	0.22	0.13	12	F
228	545	H	0.46	2	0	0.21	0.13	18	F
229	501	H	-0.29	3	0	0.20	0.13	18	F
230	546	H	0.46	2	0	0.19	0.13	41	F
231	678	I	-0.37	1	0	0.18	0.13	16	F
232	700	I	0.48	2	1	0.19	0.13	32	F
233	209	BB	-0.10	3	0	0.18	0.13	21	F
234	256	BB	0.62	1	0	0.17	0.13	39	F
235	208	BB	-0.11	1	1	0.18	0.13	14	F
236	257	BB	0.67	2	0	0.18	0.13	22	F
237	62	M	-0.26	1	0	0.17	0.13	32	F
238	106	M	0.54	4	1	0.18	0.13	23	F
239	61	M	-0.28	3	1	0.18	0.13	11	F
240	107	M	0.56	1	1	0.19	0.13	21	F

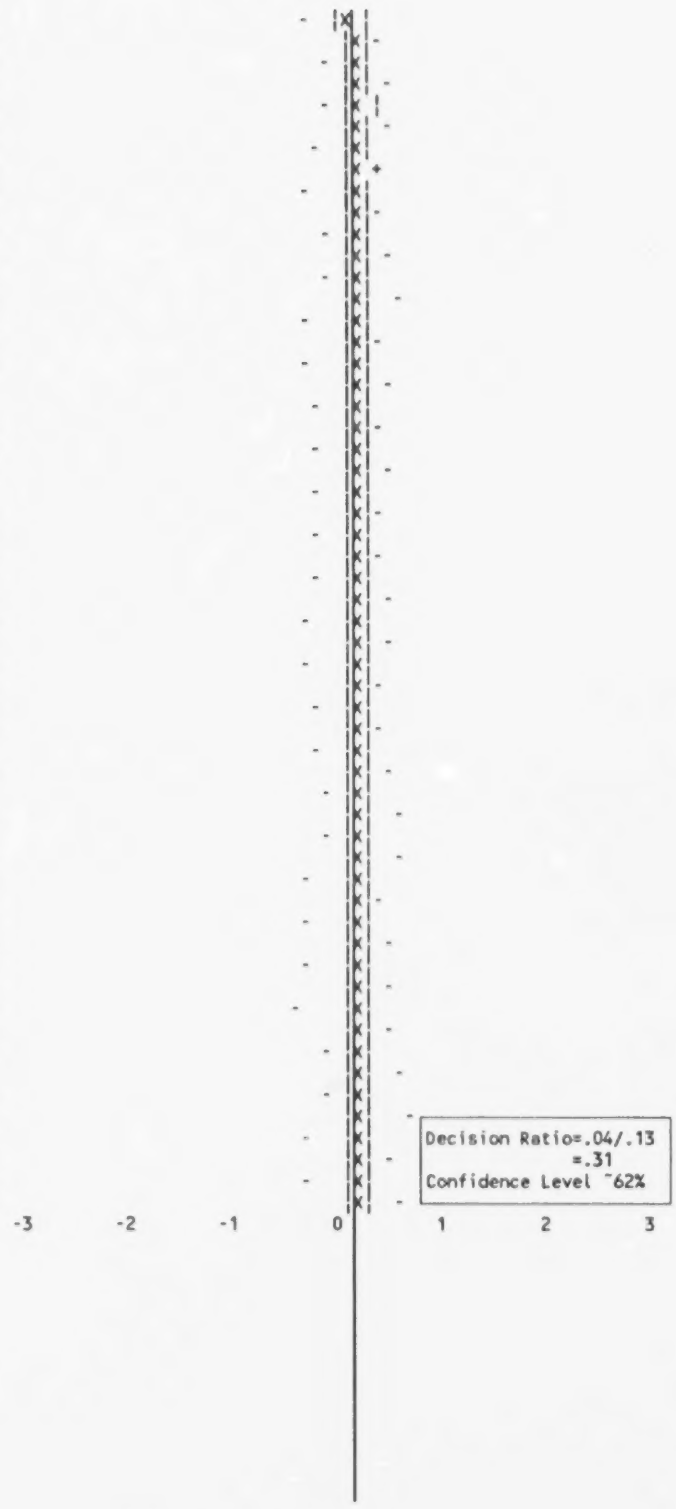


Table 1

Item Pool Distribution

Subtest	Content Specifications	Number of Items	Calibrations Mean	Std Dev	Minimum	Maximum
1	20%	147	-.06	.96	-2.89	2.36
2	20%	165	-.07	1.00	-2.21	2.94
3	20%	142	-.07	1.06	-3.61	2.97
4	20%	135	-.05	.97	-2.80	2.97
5	10%	72	.09	.97	-2.24	3.84
6	10%	65	.25	.96	-2.78	2.04
Total	100%	726	-.02	1.00	-3.61	3.84

Table 2

Comparison of Ability Measures and Test Length

	Ability Measures Mean	SD	N	Test Length Mean	SD
CAT Max 240	.23	.57	645	85	61
Cat Max 100	.23	.56	645	66	21
Long PAP	.24	.57	600	189	-
Short PAP	.22	.48	645	109	-

Figure 1

	Percent of Examinees			
	≥90 Confidence Level	80-90% Confidence Level	70-80% Confidence Level	≤70 Confidence Level
CAT Max 240	.895	.023	.031	.051
CAT Max 100	.728	.067	.082	.122
Long PAP (189 items)	.715	.090	.092	.103
Short PAP (109) items)	.583	.130	.088	.198